

序 一

统计学有很长的历史,在工农业、科技、商业、社会科学、政府决策等领域已得到普遍的应用,已经或正在成为社会发展的必不可少的工具。统计学的普遍应用只是近 40 年的事,得益于计算机的飞速发展。毫不夸张地说,没有计算机及统计软件的支撑,统计学的发展绝大部分只能停滞在理论研究上。一个新的统计方法,如果没有统计软件的支撑,是很难普及的。统计软件成为统计学发展的重要方面。

中国学者对统计软件的关注由来已久。早在“文化大革命”期间,由于地质勘探的需要,中国许多地质大队和大的油田花了大量的人力编制应用多元统计分析的统计软件。可惜编制这些软件的力量没有能够集中、提高,绝大多数团队自然衰退。以中国科学院前计算中心(现属中国科学院数学与系统科学研究院)的张建中、杨自强研究员为首的团队,一直在发展中国人自己的大型系统统计软件。在 20 世纪 80 年代,他们推出了统计软件包 SASD,对中国应用统计的发展起了很大作用。随着中国的改革开放,许多国际通用的大型统计软件包如 SAS、SPSS、MATLAB、R 等逐渐进入中国的统计教育和统计软件市场,中国人自己发展的各种统计软件包的国内市场却日益缩小。

中国是一个大国,需要有中国人自己发展的视窗系统,同时也需要具有自主知识产权的大型统计软件系统。这样的想法不是“闭关自守”,也不是“狭隘的民族主义”。

统计是一门大的学科,包括了众多的方法,如数据的描述性统计、单变量的统计分析(如 t 检验、 F 检验等)、多变量的统计分析、定性数据的数据分析、生存分析、实验的设计和建模、非参数统计等。一个大的统计软件包大多是由一个公司、一个研究单位或一所大学的一个大的团队来研制、发展、推销及服务。使我万分惊讶的是,DPS 统计软件是由唐启义教授个人用了 20 多年的时间,以“愚公移山”的精神研制和发展起来的。该软件涉及面广,使用方便,很受广大用户欢迎。唐启义教授坚持不懈的精神,令人感动。相信在唐启义教授的呵护下,DPS 统计软件会与时俱进,更加人性化,不仅在中国处处开花结果,而且能深入到全世界广大华语地区。



2009 年 5 月于中国珠海市

序 二

——青目睹人少，问路白云头

有所谓“文字的学问”，有所谓“数字的学问”。从科举制度迄今，文字学问的重要性不言而喻。相形之下，数字学问的普及与认知，尚有一段漫长的路。常言道：“数字会说话”，此言不假。但是当数字说话时，您听得懂、看得明白吗？统计学，作为研究数字学问的一门主要学科，在中国历史源远流长。相传在伏羲时代，人们便开始“结绳记事”（详见《周易》），《管子》“问篇”亦提及：“不明于数，欲举大事，如舟之无楫而欲行于大海也。”商鞅在公元前 390 年更提出：“强国需知十三数”，为统计的重要性立下经典。近来在商场上盛行的古代兵书《孙子兵法》更直言：“多算胜，少算不胜。”这些都说明统计在古代中国受重视的程度，数千年后的今天读起来依然意味深长。

计算机当属 20 世纪对人类历史推动最大的技术。自 1946 年在美国宾夕法尼亚大学诞生第一台计算机以来，计算机在功能上以飞快的速度发展，影响领域也从最开始的军事领域拓展到经济社会，以及其他各个领域。1974 年，Intel 公司发布了其面向个人机的八位微处理器芯片 8080。随着超大规模集成电路和微处理器技术的进步，计算机技术更从过去主要集中在大型机和小型机高端领域逐渐进入普通人的生活。自 20 世纪 80 年代初第一台个人计算机面世以来，计算机的成本快速下降。之后，随着互联网技术、多媒体技术的飞速进步，计算机技术快速发展。1993 年互联网开始商业化运行，1995 年 Windows 95 发布，日新月异的技术促进了计算机的快速发展与普及。1997 年 IBM 深蓝计算机战胜了人类国际象棋世界冠军，标志着人工智能发展日趋成熟。21 世纪以来，计算机发展更为迅猛，其与通信技术等的相互渗透和融合，大大推动了人类社会的进步。

计算机硬件的发展为统计学的发展提供了难得的历史机遇。统计学发展初期的方法虽然很好，但是由于计算工作量大，统计分析方法并没有在实践中有效推广开来。电子计算机的发展促进了软件的开发与商业化，推广了统计学的应用。具体来说，计算机从两个方面促进了统计学的发展。一方面，统计计算技术使得统计学家可以在计算机上验证自己的理论，推动了统计理论的发展；另一方面，大量统计软件的普及使得统计分析广泛应用到社会的各个方面，这些实践活动又推动了统计理论的发展。总之，现代统计学已经离不开计算机的发展，计算机已成为统计学发展的基本工具。

根据 1977 年 Kohm 在美国统计学会上的介绍,当时美国有 55 个公众可用的统计软件。1981 年,Francis 的著作中介绍了当时世界各国已有的统计软件 117 个。随着统计方法的不断改进,统计软件也在不断改进。70 年代百家争鸣以来,研发出了一批目前占据主流地位的统计分析软件,比如 SAS、MINITAB、BMDP、SPlus、SPSS 等。由于使用目的的不同,这些统计软件有着不同的使用人群。一般而言,BMDP 和 SPSS 比较受生物医药和社会科学领域人员的喜爱,主流的应用统计人员则更倾向于 SAS,而 SPlus 则更适合于进行统计研究的科研工作者。SAS 公司是目前全世界最大的统计软件公司。其他享有较好声誉的统计分析软件还有 SYSTAT、DataDesk 和 JMP 等。这些软件都是商业软件,由于比较昂贵,在教育界未能充分普及。20 世纪 90 年代末期,出现了类似于 S 语言的 R 软件,该软件平台是免费的,大大增加了其普及性,而由于其灵活性及强大的编程实现能力,该软件获得迅速发展和推广。

在中国统计学快速发展的过程中,统计分析软件必不可少。虽然 20 世纪 80 年代以来,各科研单位所编制的统计分析软件和程序数不胜数,但是随着统计学的发展,绝大多数统计软件由于各种各样的原因而渐渐淡出市场,使得国外较规范的统计分析软件如 SAS、SPSS、BMDP、GLIM 等占据了国内统计分析软件的大部分市场。但是,国内也产生了一批优秀的统计分析软件,如 NOSA、SPLM、Qstat、PEMS、Markway、DPS 等,其中 DPS 分析软件便是最为出色的统计分析软件之一。它立足国内市场,从初期主要用于处理农林生物科研数据起步,经过不断改进,至今已发展成为一款综合性的大型统计软件,尤其是它还开发了具有全球领先地位的均匀实验设计、混料实验设计、动态聚类分析几个功能模块的独创算法。如今,DPS 应用已遍及全国各地,涉及自然科学和社会科学各个领域,如农林牧渔、医药卫生、气象、水文地质、工程计算、市场调研、社会经济、金融等。

改革 30 年,我国经济一天比一天繁荣,但我们的教育与科研并没有明显的同步提升,距离国际水平尚有一大段发展空间。许多方面尚需努力与世界接轨,更遑论领导世界潮流。不过从另一个角度看,我们也不该一味迷信西方理论,而应该从我国国情出发,建立一套适合我国学术与教育的统计软件。这正证实了《孙子兵法》“虚实篇”所言:“致人而不致于人”。换言之,如何自力更生,建立一套属于中国人特质的统计分析软件,而不将就于外国人的思维与文化,进而进军国际市场,影响全世界。作为国内最为优秀的统计分析软件之一,DPS 统计软件研发始于 1988 年,经过 20 多年的努力,发展到现在的规模,填补了国内具有自主知识产权的大型统计软件的空白。

统计大师赵民德先生在《探索真相》一书的导读中引用了《鹿鼎记》第二十二回中韦小宝为了学“一指禅”就询于少林寺首座澄观和尚的一段对话。此话给我印象极深,特别引用如下:

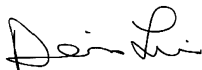
“韦小宝：‘你说那一指禅并不难学，可是从少林长拳练起，一路路拳法练将下来，练成这一指禅，要几年工夫？’

澄观道：‘这在般若堂的典籍中是有记载的。五代后晋年间，本寺有一位法慧禅师，生有宿慧，入寺不过三十六年，就练成了一指禅，进展神速，前无古人，后无来者。料想他前生一定是一位武学大宗师，许多功夫是前生带来的。其次是南宋建炎年间，有一位灵兴禅师，也不过花了三十九年时光。那都是天纵聪明、百年难遇的奇才，令人好生佩服。前辈典型，后人也只有神驰想象了。’”

统计学之深奥，多少人穷极一生都无法参透，吾等凡夫俗子又如何在有生之年得到真谛呢？《DPS[®]数据处理系统》让我们看到了一线曙光。该书内容涵盖之广，仿佛一本统计百科全书。全书8篇52章，包山包海，几乎包括了目前统计应用的主要方向。全书由浅入深，精辟易懂，配合软件操作，让一门高深的统计学可以在短期内得以理解并熟练地应用在生活周边的事务上。那统计学的“一指禅”在一年半载内借由该书得以学到六七成功力，不亦乐乎！

浙江大学唐启义教授是个认真的学者，同时也是热情而务实的工作者。从另一个角度看，唐教授是个理论扎实的统计学家，同时也是应用统计实务的执行者。更者，他是个计算机专家。通过他精心设计，妙笔生辉，该书充分反映了这些特质，也适时地满足了我国目前统计界的需求。

是以为序。

林共进 

2009年春写于美国宾夕法尼亚州

第 6 版前言

DPS 数据处理系统经过 30 多年的开发升级以及数以万计用户的应用反馈,目前已具备了较为完整、适用于各个专业试验设计及试验结果数据统计分析的能力。同时,DPS 数据处理系统已经可以在 Windows 系统及不同芯片架构的国产操作系统下运行。

DPS 数据处理系统反映了我国在具自主知识产权的统计分析软件开发方面实现核心技术的自主可控,解决了数据统计分析等关键领域长期依赖国外技术的问题,实现了各行各业数据统计分析的国产替代。

《DPS[®]数据处理系统》是基于我们开发的 DPS 数据处理系统计算机软件,介绍 DPS 数据处理系统软件试验设计及数据统计分析的工具书。目前版本是第 6 版,篇章结构仍沿用第 5 版。第 6 版主要新增内容有:

第一卷,完善了广义线性模型功能,将复杂的广义线性模型功能以方便操作的图形界面的形式提供给读者。另外,在第一卷中,作者增加了多指标 Kendall 趋势检验等功能。

第二卷,在 BP 神经网络建模中新增了目前常用的激励函数,以及对 BP 神经网络模型中各输入变量对输出变量的影响效应评估的改进。

第三卷,增加了大量的统计内容。这里介绍了作者在综合评价、多属性决策及机器学习中高维数据降维到一维方面的研究成果,即作者提出的最大熵-最小残差非线性数学模型。该数学模型解决了令人头疼的多指标综合评价、多属性决策过程中各个指标权重系数如何科学、客观、精确估计的世界性难题。在专业统计中,还增加了测量系统分析、统计质量控制、多因子分析的决策试验及评价实验室法(Dematel 法)和解释结构模型法(ISM)等内容,完善了 Meta 综合评价分析方法等。同时根据用户建议,对专业统计分析中一些统计功能结果输出进行了完善和充实。

本书各卷和前几版一样,用户界面中对于图中颜色的表述,均是 DPS 软件使用时在屏幕上的显示特征,和书中黑白印刷示意图有所差别,读者在阅读本书时需予以注意。

包括在国产操作系统下,各个版本的 DPS 数据处理系统软件试用版以及相关的帮助文档可从 DPS 数据处理系统网站(网址: www.dpsw.cn)下载。DPS 高级版下载安装之后在安装文件夹下面有一个“例子数据(高级版)”,里面有本书所附的所有例子数据,以供用户了解、练习 DPS 数据处理系统的功能。

DPS 数据处理系统既是我们 30 多年不懈努力的结晶,更是与广大 DPS 统计软件用户一如既往地反馈、支持密不可分。用户的需求是我们软件开发的动力!在此,我们对广大用户的使用、反馈致以衷心感谢。

本书再版得到了农业农村部“农作物病虫害监测、预警与防治服务”(代码: A120301)和杭州睿丰信息技术有限公司的资助。本书成稿过程中得到了作者所在单位和科学出版社的大力支持。这里作者感谢浙江大学生命科学学院冯明光教授和中国水稻研究所胡国文研究员多年来对作者工作的支持和帮助。对已故的浙江大学原副校长程家安教授、湖南省农业厅李绍石研究员、华南农业大学庞雄飞院士在作者从事科研的过程中给予的殷切关怀和鼓励致以衷心的感谢。同时对 30 多年来对作者在 DPS 软件开发、文稿撰写过程中给予帮助的各位朋友、同事、学生,以及我的家人,一一表示感谢。

最后,对 DPS 数据处理系统的广大用户以及关心和扶持 DPS 成长的各个领域的专家、学者表示衷心感谢。

唐启义

2025 年 11 月

第 5 版前言

这是一本介绍试验设计及数据统计分析的工具书,是基于我们开发的 DPS 数据处理系统计算机软件而成的。DPS 数据处理系统经过我们 30 多年的开发以及数以万计的用户的反馈、升级,已较好地满足各个专业试验设计及试验结果数据统计分析的需求。本书也反映了我们在具有自主知识产权的统计分析软件开发方面几十年来不懈努力的成果。

目前的第 5 版,篇章结构仍沿用第 4 版,主要新增内容有:

第一卷,改进了随机区组试验每个处理标准差、标准误的计算方法,即将区组效应去掉之后再计算;增加了两随机变量回归分析、Joinpoint 回归(折线回归)分析功能。

第二卷,这次再版主要新增了两个章节,即第 16 章“高维数据 Lasso 回归分析”和第 20 章“有序约束下的多峰识别模型”。前者既提供了高维数据在变量数量大(多重共线性)且大于样本数时的回归建模技术,又提供了自变量含有分类变量且有缺失数据情形下筛选重要因子、建立回归模型的途径。后者介绍了作者首次提出的有维度限制的多峰识别模型,它提供了解决一维、二维、三维及更高维有序样本的统计分析(如聚类分析)的技术。另外在时间序列分析章节中增加了向量自回归模型、格兰杰因果检验、协整检验和误差修正模型。并对 BP 神经网络算法进行了改进,使之可适合于含缺失数据的情形,并在 BP 神经网络模型中增加了估计各输入变量对输出变量的影响效应的功能。

第三卷,这次再版将面板数据一章调整到第二卷。功能上主要新增了水文气象频率分析,其中对 Wakeby 分布的参数估计提出了独创的新的算法,使拟合精度提高,进行百年一遇、千年一遇重现期估计时更精准。同时根据用户建议,对专业统计分析中一些统计功能结果输出进行了完善和充实。

本书各卷,和前几版一样,用户界面中对于图中颜色的表述,均是 DPS 软件使用时在屏幕上的显示特征。因此和书中黑白印刷示意图有所差别,读者在阅读本书时需予以注意。

DPS 数据处理系统软件试用版以及相关的帮助文档可从 DPS 数据处理系统网站(网址: www.dpsw.cn)中的“下载中心”下载。DPS 高级版下载安装之后在安装文件夹下面有一个“例子数据(高级版)”,里面有本书所附的所有例子数据,以供用户了解、练习 DPS 数据处理系统的功能。

DPS 数据处理系统既是我们 30 多年来不懈努力的结晶,更是广大 DPS 统计

软件用户一如既往地支持的成果。读者和用户的需求是我们软件开发的动力!

本书再版得到了国家重点研发计划项目(项目编号:2016YFD0201302)的资助。作者所在单位和科学出版社给予了大力支持。这里作者感谢浙江大学生命科学学院冯明光教授多年来对作者工作的支持和帮助。对浙江大学原副校长程家安教授、中国水稻研究所胡国文研究员及已故的湖南省农业厅李绍石研究员、华南农业大学庞雄飞院士对作者在 DPS 统计软件开发、升级,以及文稿写作过程中给予的殷切关怀和鼓励致以衷心的感谢。同时特别感谢郭若艳同志在这次再版过程中为校对文稿所付出的大量心血。最后对方开泰教授、王静龙教授、周仁郁教授、刘明芝教授、方萍教授、黄敬峰教授、张承恩高级工程师、董玉恒教授及夏结来博士等给予的帮助表示感谢。多年来浙江大学研究生高薇、孙传恒、唐洁、沈爱华、时嵩、彭奇、宛庭利、刘顺、罗巍等在协助文稿整理、图表编辑处理方面做了很多工作;浙江大学昆虫科学研究所、浙江大学环境毒理研究所的同事给作者工作以很大帮助,在此一并表示感谢。

最后,对 DPS 广大用户以及关心和扶持 DPS 成长的各个领域的专家、学者表示衷心感谢。

唐启义

2020 年春于浙江大学紫金港校区

第 4 版前言

光阴荏苒。DPS[®]数据处理系统软件及其配套专著自 1997 年首次出版以来,经过近 20 年不懈努力,DPS[®]数据处理系统软件得到了长足的发展。配套专著先后以《实用统计分析及其计算机处理平台》《实用统计分析及其 DPS 数据处理系统》《DPS[®]数据处理系统——实验设计、统计分析及数据挖掘》,以及三卷本的《DPS[®]数据处理系统》配套专著形式出版,其篇幅亦从 60 多万字增加到目前的近 200 万字。

目前的第 4 版,篇章结构仍沿用第 3 版,主要的新增内容有:

第一卷,在相关和回归分析(第 6 章)中增加了 Deming 回归、具重复测量试验回归;在分类数据模型分析(第 8 章)中增加了广义线性模型;非参数统计(第 9 章)增加了 Passing-Bablok 回归;混料试验设计与分析(第 12 章)增加了极端顶点设计功能。另外,在第 2 章、第 4 章和第 5 章改进了数据分析方法、用户界面,增加了各个处理直方图比较,以及模型残差作图分析。

第二卷中,作者对相关、回归分析章节内容按是否是多个因变量进行了重组,将典型相关分析、多重筛选逐步回归、偏最小二乘回归分析及线性联立方程合并成“多因变量统计分析”一章,并将原来结构方程模型与路径分析放到本章节。统计分析功能方面,作者改进了逐步回归分析和双重筛选逐步回归分析的用户界面,优化了双重筛选逐步回归相关算法,并在时间序列分析章节中新增了单位根检验等内容。

第三卷中的专业试验统计,增加了面板数据统计分析、多点试验统计分析、格子设计统计分析等方法。并根据用户建议,对一些统计功能结果输出进行了完善、充实,改进许多 DPS 使用时的用户界面。

本书各卷中,对于图中颜色的表述,均是 DPS 软件使用时在屏幕上的显示特征。因此和书中黑白印刷示意图有所差别,读者在阅读本书时需予以注意。

DPS 数据处理系统软件试用版以及相关的帮助文档,可从 DPS 数据处理系统网站(网址:<http://www.dpsw.cn>)中的“下载中心”下载。注意 DPS 数据处理系统有高级版(最新版本)和标准版(2007 年底以前)两个版本。DPS 系统高级版下载安装后,在安装文件夹下面有一个“例子数据(高级版)”文件夹,其中有本书所附的所有例子数据,供用户了解、练习 DPS 数据处理系统及其使用。

在广大读者和 DPS 统计软件用户支持下,经过这三年多的努力,DPS 系统功能又得到了进一步的增强。作者衷心希望广大读者一如既往地支持 DPS 这一国

产统计分析软件,读者和用户的需求就是我们软件开发的动力!读者和用户在实验设计、统计分析和数据挖掘方面如遇到困难,可以通过访问我们为 DPS 提供的专门网站 <http://www.dpsw.cn>,即可得到技术咨询和软件升级。

本书再版得到了国家重点研发计划项目(2016YFD0201300)的资助。作者所在单位和科学出版社给予的大力支持。这里,作者感谢浙江大学生命科学学院冯明光教授多年来对作者工作的支持和帮助。对浙江大学原副校长程家安教授、湖南省农业厅李绍石研究员、中国水稻研究所胡国文研究员及已故的中国科学院庞雄飞院士对作者在 DPS 统计软件开发、升级,以及本书写作过程中给予的殷切关怀和鼓励致以衷心的感谢。并对方开泰教授、王静龙教授、周仁郁教授、刘明芝教授、方萍教授、黄敬峰教授、张承恩高级工程师、董玉恒教授及夏结来博士等给予的帮助表示感谢。多年来浙江大学研究生高薇、孙传恒、唐洁、沈爱华、时嵩、彭奇、宛庭利、刘顺、罗巍等在协助文稿整理、图表编辑处理方面做了很多工作;浙江大学昆虫科学研究所、浙江大学环境毒理研究所的同事给作者的工作以很大帮助,在此一并表示感谢。

最后,对 DPS 广大用户以及关心和扶持 DPS 成长的各个领域的专家、学者表示衷心感谢。

唐启义

2016 年夏于浙大紫金港

第 3 版前言

《DPS[®]数据处理系统》(第 3 版)是作者基于自主开发的 DPS 统计软件而撰写的,以实用为目的的统计工具书。

随着 DPS 统计软件功能的不断完善,相应配套专著已变得越来越厚。许多 DPS 用户、读者反映书本太厚了一些,使用起来不甚方便。因此,这次出版对版面进行了较大调整,并将整个书稿分成三卷出版,第一卷为基础统计及实验设计;第二卷为现代统计及数据挖掘;第三卷为专业统计及其他。

第一卷包括了第二版的第一篇和第二篇及原来“抽样技术”一章,并将原“概率统计模型”一章中的加权列联表分析、多因子综合相关分析放在本卷第 8 章“分类数据模型分析”里面。

在第一卷中,新增了相关和回归分析(第 6 章),系统介绍了相关和回归分析原理、应用及应用中需注意的一些问题。另外,在第 1 章里面增加了“工作表中图形输出”和“DPS 系统应用常见问题解答”两节,在第 2 章里面增加了“缺失值处理”,在第 7 章里面增加了“Poisson 分布抽样下多样本检验”,在第 11 章“多因子优化设计与分析”里面,改写了实验设计工作界面,增加了区组设计的统计分析内容。

第二卷主要包括了第二版的第三篇(多元统计分析)、第四篇(数学模型模拟分析)和第七篇(时间序列分析)以及第八篇中的第 43 章“神经网络和支持向量机”。

在第二卷中新增的内容主要是第 19 章“结构方程模型”,介绍了如何在 DPS 系统下,借助良好的用户界面,实现结构方程的建模与分析。另外,在第 15 章里面增加了“线性联立方程”一节,并以第 2 版第 43 章“神经网络和支持向量机”为主体,增加了“随机森林”,将该章命名为“数据挖掘”。在第 3 版的第 21 章“回归方程模型”主要是原第 2 版的第 30 章“非线性回归模型”内容,但充实了许多内容,如联立方程模型功能得到加强。在时间序列分析领域,增加了时间序列数据的游程检验、谐波分析、奇异谱分析等功能。

第三卷主体内容为第 2 版的第三篇“专业实验统计”、第六篇“常用数值分析”和第八篇“其他数据分析方法”。

在专业实验统计中,这次我们主要增加了生物测定及药物代谢统计学分析、品种区试 AMMI 模型分析作图技术,并根据用户建议,对一些统计功能结果输出进行了完善、充实,修改了许多 DPS 使用时的用户界面。

在本书各卷中,对于图中颜色的表述,均是 DPS 软件使用时在屏幕上显示特征的描述。因此和书中黑白印刷的示意图有所差别,请读者在阅读本书时需予以

注意。

在广大读者和 DPS 统计软件用户的支持下,经过作者三年多的努力,DPS 系统功能又得到了进一步增强。希望广大读者一如既往地继续支持 DPS 这一国产统计分析软件,读者和用户的需求就是我们软件开发的动力!读者和用户在实验设计、统计分析和数据挖掘方面如遇到困难,可以通过访问我们为 DPS 提供的专门网站 <http://www.dpsw.cn> 得到技术咨询和软件升级。

本书这次再版得到了国家重点基础研究发展计划(“973”)课题(编号:2010CB126201)部分资助。作者所在单位和科学出版社给予了大力支持。这里,作者感谢浙江大学生命科学学院冯明光教授多年来对作者工作的支持和帮助。对浙江大学原副校长程家安教授、湖南省农业厅李绍石研究员、中国水稻研究所胡国文研究员及已故的中国科学院庞雄飞院士对作者在 DPS 统计软件开发、升级,以及本书写作过程中给予的殷切关怀和鼓励致以衷心的感谢。对方开泰教授、王静龙教授、周仁郁教授、刘明芝教授、方萍教授、黄敬峰教授、张承恩高级工程师、董玉恒教授及夏结来博士等给予的帮助表示感谢。浙江大学 2004 级软件工程专业唐睿和外语系沈宁宁协助完成了 DPS 数据处理系统软件英文版的翻译与转换工作;研究生宛庭利、刘顺、罗巍,湖南农业大学实习生徐西林、刘云协助第 3 版书稿里面的图表编辑处理;浙江大学昆虫科学研究所、浙江大学环境毒理研究所的同事给作者的工作以很大帮助,在此一并表示感谢。

最后,对 DPS 广大用户以及关心和扶持 DPS 成长的各个领域的专家、学者表示衷心感谢。

唐启义

2013 年春于浙大紫金港

第 2 版前言

《DPS[®]数据处理系统——实验设计、统计分析及数据挖掘》(第 2 版)是作者基于自主开发的 DPS 软件,以实用为目的的统计工具书。

根据读者、用户建议,这次再版除对原有的一些内容进行修改、补充外,还增加了作者在过去两年研发出来的若干新功能的介绍。全书新增了 3 章共 30 多节,并对一些章节的顺序进行了调整。新增的 3 章是:

第 11 章“混料实验设计与分析”。在第 1 版的均匀实验设计技术取得的重大进展的基础上,本章介绍了自主开发的、较目前已有的混料实验设计方法更有效的混料实验设计的新技术,该技术能适合各种情形的混料设计,特别是组分多、实验次数大、限制条件“苛刻”的情形。同时提供了混料实验数据分析和优化的数学模型,为工业、农业和科学实验中应用混料设计提供了新的技术。

第 12 章“数据包络分析与随机前沿面分析”。主要用来评价相同类型部门或单位(称为决策单元)中各成员之间的相对有效性,其中包括非参数的数据包络分析(data envelopment analysis, DEA)和属于参数方法的随机前沿面分析(stochastic frontier analysis, SFA),并给出了在实际工作中计算生产效率时应用较多的 Malmquist 指数。

第 16 章“序贯实验分析”。在进行下一次实验以前,已经知道前面实验的结果,这种一次接着一次的抽样实验为序贯实验。本章提供了在序贯实验中计算各接收时刻的接收概率和各时间区段的拒收概率的方法,并给出了工作特性曲线。

除这 3 章外,还在方差分析、非参数检验、多因子优化与分析、量表分析和顾客满意度指数模型、多变量统计检验、回归分析、多因子分析、非线性回归模型、时间序列及多指标综合评价等 10 多个章节,增加了新的统计分析功能,如实验设计和方差分析中的 Box-Behnken 设计、平衡不完全区组设计实验方差分析、增广随机区组设计及方差分析、多元方差分析的线性模型方法,回归分析中的稳健回归、积分回归和主导分析,多变量时间序列的独立分量分析,以及结合分析、展开法、非参数回归、支持向量机和投影寻踪分类等。

经过这两年多的努力,尽管 DPS 系统功能又进一步增强,但距离广大读者的要求仍差距甚大。希望广大读者一如既往地继续支持 DPS 这一国产统计分析软件,及时指出目前版本的缺点和不足,以便我们继续改进,加速开发新的数据分析技术。读者可通过访问网站 <http://www.chinadps.net> 得到技术咨询和软件升级。

本书再版得到了国家高技术研究发展计划(863 计划)项目(2006AA10Z217)和农业公益性行业科研专项(200803003)的资助。在本书出版过程中,作者所在单位和科学出版社给予了大力支持。这里,作者特别感谢浙江大学生命科学学院冯明光教授多年来对作者工作的支持和帮助。对浙江大学原副校长程家安教授、湖南省农业厅李绍石研究员、中国水稻研究所胡国文研究员及已故的中国科学院庞雄飞院士对作者在 DPS 统计软件开发、升级,以及本书写作过程中给予了殷切关怀和鼓励致以衷心的感谢。并对方开泰教授、王静龙教授、周仁郁教授、刘明芝教授、张承恩高级工程师、董玉恒教授及夏结来博士等给予的帮助表示感谢。浙江大学软件工程专业唐睿和外语系沈宁宁协助完成了 DPS 数据处理系统软件英文版的翻译与转换工作;研究生时嵩、彭奇、高薇、缪清玲协助完成了本书文稿的文字处理;浙江大学植物保护系、昆虫科学研究所农业部作物病虫分子生物学重点开放实验室以及浙江省农业遥感与信息技术重点研究实验室的同事给作者的工作以很大帮助,在此一并表示感谢。

最后,对 DPS 广大用户以及关心和扶持 DPS 成长的各个领域的专家、学者表示衷心感谢。

唐启义

2009 年春节于杭州华家池

第 1 版前言

呈现在读者面前的《DPS[®] 数据处理系统——实验设计、统计分析及数据挖掘》是作者于 2002 年出版的《实用统计分析及其 DPS 数据处理系统》的补充和发展。和原来版本一样,它既是一本以实用为目的的统计工具书,更是一份请柬。它邀请读者进入 DPS 数据处理系统漫游,寓复杂繁冗而又沉闷乏味的统计分析于轻松自如和无穷乐趣之中。

现代科学研究,不管是自然科学还是社会科学的各个领域(物理、化学、工程、医学、生物、经济、社会等),只要存在带有随机性的数据,就需要应用数理统计方法去有效地收集、整理和分析,以便得出正确的结论,为相关决策行动提供科学依据。

统计学的形成可以追溯到 20 世纪 20 年代。当时 Fisher 等以农业科学实验为对象,探讨实验资料的整理和统计分析方法。1922 年,Hayes 研究亲代与子代蛋白质含量的相关性,提出了相关分析的理论。翌年,Fisher 和 Maekenzie 研究肥料对马铃薯产量的影响,首次提出了方差分析和交互作用的概念。1925 年,Engel-dow 研究不同品种在不同地区和不同年份的产量表现时,进一步阐述了因素间的交互作用。Fisher 等正是在研究农业科学实验方法的基础上,建立起了统计学基本理论和方法。这些已被广泛地应用于科学研究、工农业生产和现代经济管理。

随着经典数理统计科学的发展,在 20 世纪四五十年代,多元统计分析在 Pearson 及 Fisher 等人开拓性工作基础之上又有了坚实的理论基础,如关于多元正态分布的理论著述、Anderson(1958)的《多元统计分析导论》等著作,标志着统计学进入到一个新的发展阶段。但是,多元统计分析的发展、完善和广泛应用,离不开计算机科学和计算数学的发展。由于多元统计分析的计算工作量十分惊人,没有计算机是不可想象的。正是由于计算机技术的飞速发展,一些独特的多元统计分析技术,如偏最小二乘法、主成分分析、典型相关(又称“典范相关”)、因子分析、对应分析、判别分析、聚类分析等,在农林、医药、气象、地质、生物、心理、工业、经济、社会等许多研究和生产领域获得了广泛应用。目前,多元统计分析已是统计学中最活跃的分支学科之一。

传统的自然科学和社会科学研究以定性描述为主。Fisher 等创立的一套经典统计方法在定量研究方面迈出了一大步。但囿于其固有缺陷,实验结果的统计推断不能对有关数量关系从机理上予以合理解释,因而其统计模型仍属经验范畴。因此,现代科学工作者们正力图使具体学科的理论及研究对象同数学有机地结合起来,以模型来描述、解释自然现象和社会、经济的运动规律,从而解决科研和生产

中的实际问题。

自 20 世纪六七十年代以来,一些新的统计理论和分析技术发展起来。如 20 世纪 60 年代 Zadeh 提出的模糊集合论在各个领域的应用已十分广泛。实践证明,模糊数学方法在农业、图像识别、天气预报、地质地震、交通运输、医疗诊断、信息控制、人工智能等领域的应用已初见成效,其发展前景广阔。作为预报和控制的有力工具,研究系统变化规律的时间序列分析技术目前正被广泛应用于地质、石油、气象、商业、工程控制等领域,在农业中已被用于作物病虫害的中长期预测预报,并获得了可喜的成功。灰色系统是我国学者邓聚龙先生于 20 世纪 80 年代初提出的用于控制和预测的新技术,目前亦在我国农业、社会和经济等领域广泛应用,并已取得显著成就。

统计学的发展源远流长,内容十分丰富。迄今为止,国内外统计学专著和教科书已有多种版本见诸于世。如果按照一般统计学著述的内容和体裁,我们没有必要再推出一本教科书似的理论专著,让广大科技工作者在复杂的数学公式堆中艰难跋涉。本书的编写宗旨强调实用性。作者在内容上刻意求新,形式上力求通俗易懂,在取材上尽可能全面,并注意反映国内外在实验设计和统计分析方面的最新研究进展。书中介绍的 400 多种统计分析方法,可以满足各个领域的统计分析需要,且都能在作者设计制作的 DPS 计算机数据处理系统中随意调用,自如操作。本着服务于读者、服务于用户的最高宗旨,作者对入选的 400 多种统计分析方法分门别类,每一类(种)统计分析技术按照方法或原理简介与操作示例两个基本部分编写,全书共分 8 篇 41 章。同时,作者按书中篇目类别设计 DPS 系统的套叠下拉式主菜单和子菜单,将带给读者和用户极大的便利。下面分篇给予简单介绍,以引导读者高效率地使用书中的各种分析技术和 DPS 数据处理系统。

第一篇主要介绍 DPS 系统的性能、特点和操作要领。DPS 数据处理系统是作者独立研究开发的大型通用多功能数据处理分析应用软件系统。该系统在 IBM PC 及其兼容机、中文 Windows 98/2000/XP 下运行。因此,在使用 DPS 系统之前,建议读者首先浏览本篇内容,熟悉该系统将使读者在数据处理和统计分析中事半功倍。

第二篇的编写充分考虑到经典实验统计的完整性和系统性,提供了比较全面的实验设计和实验结果统计分析的内容,如 t 测验、方差分析、列联表分析、非参数检验、最优回归实验设计与分析以及多元方差分析等,尤其对方差分析给予了较为详尽的介绍。因此,各个层次、各个领域的读者都可使用这些基本的统计学分析方法。

第三篇介绍了一些应用于某些较特殊专业领域的统计分析技术,如心理统计中的量表分析、顾客满意指数的结构方程模型分析、实验评价的 ROC 曲线分析、地理统计分析,还为育种学工作者提供了品种区域实验的统计分析技术,为生态学

工作者提供了动植物种群数量抽样、动植物种群数量空间分布研究、生物群落的消长演替过程分析、生存分析等方面的内容;对从事生物测定研究的科技人员来说,作者既提供了经典 Finney 概率分析方法,还重点引入近几年才逐渐发展起来的更为完善的分析工具,即时间—剂量—死亡率模型(TDM)分析技术。

第四篇以较大篇幅介绍多变量统计分析技术,并注意反映目前国内外的最新研究动态,将最新的技术提供给读者。近年来,多变量统计分析技术已被广泛应用于各个领域的科学实验中,因为人们越来越多地将研究对象作为一个有机结合的整体即多元素系统(亚系统)来对待,而不是分割开来作为单独事物处理。“从树木到森林”,抓住系统运动变化的本质,舍弃次要因素,通过简化系统的结构,探索系统运动变化的总趋势,为人类征服自然、改造自然服务。因此,偏最小二乘回归分析、主成分分析、因子分析、对应分析、典型相关分析、聚类与判别分析以及马尔可夫链分析等多因子分析技术备受自然科学和社会科学各领域科研人员的重视。在 DPS 中还提供了我们研发的含有定性变量的逐步回归分析技术。

第五篇和第六篇向读者介绍相当灵活的通用数学模型的模拟分析和数值计算工具。因前面介绍的分析方法和程式基本上都有完整的体系和要求,读者必须按部就班地照着去做。而利用第五、六篇的通用工具,读者将尽可能地发挥自己的专长和展现自己的学识水平,实现思维活动的飞跃。例如,模拟作物生长发育过程或经济活动的演替规律,建立结合自己专业和学科特点的数学模型,而不是套用别人的现成模型。读者可以借助 DPS 系统强大的数学建模功能,在计算机上实现自己的思维活动,获得所想即所见、所见即所得的效果。

第七篇介绍时间序列分析技术。时间序列分析是根据系统观测得到的时间序列数据,通过曲线拟合和参数估计来建立数学模型的理论和方法,常用于国民经济宏观控制、区域综合发展规划、企业经营管理、市场潜量预测、气象预报、水文预报、地震前兆预报、农作物病虫害灾害预报、环境污染控制、生态平衡、天文学和海洋学等方面。

第八篇为读者提供一些不属于经典统计学但近年来发展迅速且应用已经相当广泛的模型分析技术,即模糊数学分析、灰色系统分析和神经网络模型技术。只要读者有兴趣,DPS 系统可以帮助读者熟悉和掌握上述各项分析技术,并应用到研究实践中去。

综上所述,本书和 DPS 数据处理系统是密不可分的整体。作为通用的数据处理和统计分析的工作平台,作者在系统的设计方面注重专用性与通用性以及“傻瓜”性和灵活性的关系,以适应不同层次读者的需要。

专用性和通用性。DPS 的有些功能是专为某些领域或完成某一类工作而专门设计的,如“生物测定”是专为从事化学或生物制剂的生物活性测定的人员而设计的;种群空间分布型及群落参数统计分析、品种区域实验等章节都是服务于某个领

域的科学工作者。相对来说,它们的用途比较专一。但是,本书中绝大部分统计分析和建模技术都是通用的,例如丰富的数学函数、交互方式的矩阵运算功能以及通用数学模型的模拟分析技术。DPS 处理数学公式的功能非常独特,有利于读者发挥潜在的想象力和创造力。同时,数学计算功能,特别是矩阵交互式处理功能,在当前微机普及的情况下,将有助于高等院校数理统计课程的教学,使繁冗无味的矩阵运算得以在计算机上精确而快速地进行,并且在一定程度上提高学生的学习积极性和教学质量。

“傻瓜”性和灵活性。在 DPS 平台上进行经典统计分析的过程十分简单。它无需读者懂得数学推导,无需读者了解运算细节,只需读者将自己的实验数据输入电子表格,将它们定义成数据块(移动鼠标将待分析的数据“框”起来),然后调用分析功能(用鼠标点选菜单),按回车键就可在当前光标下面得到分析结果。虽然有些较复杂的统计分析需要读者输入参数,但系统会自动给以提示,如用户事先没有相关的经验,系统会给出缺省值,用户只管用鼠标点击“确定”按钮,就可以获得较优的分析结果;又如时间序列分析,系统以直观的图形方式让读者确定参数,只要读者稍有常识,无需牢记一大堆统计分析过程中的参数辨识准则,就可成为高水平的建模能手;再如在系统聚类分析中,聚类结果一次成“像”,给出的是非常直观的聚类谱系图等。系统的这些功能,好比一只傻瓜型照相机,使用时只需按快门。另一方面,DPS 系统在数据处理和模型模拟中又比较灵活,完全取决于读者的经验和知识水平。如数据行列的转换功能、通用方差分析的广义线性模型的应用、自定义数学模型的拟合、逐步回归及逐步判别分析中 F 检验临界值的随意调整等等,使读者大有用武之地,挥洒自如。

应当指出,本书不是一本基础理论教材,而是一本不需要高深数理统计知识即可看懂的统计方法论专著。事实上,要把本书中每种统计方法的理论背景在这样一本书中完全交代清楚是不太可能的,因为有些章节本身就是一本书的内容。作者的主要目的是要把读者从繁琐的统计分析中解放出来,既要用最适合、最先进的统计分析技术为自己的研究工作服务,又不被它们复杂的数学公式和运算所困扰。如果本书及其 DPS 数据处理系统能使读者在数据处理和统计分析方面有所获益或帮助,作者的基本目的就达到了。

DPS 统计软件相关图书自 1997 年中国农业出版社第一次出版以来,已经历了 10 个不平凡的年头。在这 10 年中,国内外统计软件品种繁多,竞争十分激烈,但由于有广大用户的大力支持,DPS 终于突破重围,在国内应用统计软件领域占有了一定地位。为此,我们对 DPS 广大用户以及关心和扶持 DPS 成长的各个领域的专家、学者表示衷心的感谢!

经过 4 年的努力,尽管 DPS 系统的功能又进一步增强,但离广大读者的要求还有较大差距。作者希望广大读者继续支持 DPS 数据处理系统这一国产统计分

析软件,使之不断完善。在使用过程中如发现缺点和不足请及时反馈给作者,以便再次修订时改进。作者将不辜负广大读者的期望,努力工作,不断充实新内容。为方便广大读者、用户使用 DPS 数据处理系统,及时解答用户在应用中的技术问题,加快更新 DPS 软件步伐,使之日臻完善,特设立了为 DPS 统计软件提供技术支持的网站 <http://www.chinadps.net> 和 <http://www.statforum.com>。读者可通过访问该网站及时得到 DPS 使用技术咨询和软件的升级服务。

本书的出版得到国家自然科学基金(30370914)和国家 863 计划项目(2002AA243041)的资助。在本书修订和出版过程中,作者所在单位和科学出版社给予了大力支持;浙江大学原副校长程家安教授、湖南省农业厅李绍石研究员、中国水稻研究所胡国文研究员及已故的中国科学院庞雄飞院士对作者在 DPS 统计软件开发、升级,以及本书写作过程中给予了殷切关怀和鼓励;香港浸会大学方开泰教授、华东师范大学王静龙教授、成都中医药大学周仁郁教授、湖南中医学院刘明芝教授以及第四军医大学夏结来博士等给予大力帮助;研究生孙传恒、唐洁和沈爱华协助本书文稿的文字处理;浙江大学昆虫科学研究所的各位同事给作者工作以很大帮助,作者在此一并表示感谢。

作 者

2006 年 3 月于杭州

目 录

序一

序二

第 6 版前言

第 5 版前言

第 4 版前言

第 3 版前言

第 2 版前言

第 1 版前言

第一篇 DPS[®]数据处理系统

第 1 章 DPS 系统简介	3
1.1 系统功能简介	3
1.2 DPS 系统的不同版本	6
1.3 系统运行环境与安装、使用	6
1.4 国产操作系统下 DPS 数据处理系统安装与使用	8
1.5 DPS 数据处理系统用户工作界面	9
1.6 DPS 数据文件	12
1.7 DPS 的基本操作	13
1.8 DPS 系统功能的用户定制	15
1.9 文本数值转换及字符串数值转换	17
1.10 数据行列转换及行列重排	17
1.11 分类变量的取值和编码	18
1.12 常用数据转换及多项式生成	19
1.13 数据统计分析及其建模基本步骤	21
1.14 工作表中图形的输出	23
1.15 DPS 系统函数应用	23
1.16 DPS 系统应用常见问题解答	27
参考文献	29
第 2 章 数据统计基础概念	30
2.1 数据基本参数计算及数据转换	30

2.2 常用统计分布及 DPS 统计函数	36
2.3 连续变量统计及分布的适合度检验	40
2.4 离散变量统计及分布的适合度检验	47
2.5 均值及方差的稳健估计	52
2.6 置信区间及参考值范围计算	55
2.7 混合分布参数估计	59
2.8 Pearson-III 型分布	64
2.9 异常值检验	67
2.10 缺失值的处理	70
2.11 图表处理	71
参考文献	76

第二篇 实验统计分析

第 3 章 一组样本和两组样本统计检验	81
3.1 显著性检验基本原理	81
3.2 平均数和总体差异检验	83
3.3 总体均值样本量估计	85
3.4 样本率和总体率的比较	86
3.5 Poisson 分布的均数和总体比较	88
3.6 两组样本均值差异 t 检验	89
3.7 小样本均值差异 Fisher 非参数检验	93
3.8 Bonferroni 检验	94
3.9 两组样本率差别检验	96
3.10 两总体检验样本含量及功效估计	100
3.11 概率模型拟合优度检验	103
参考文献	105
第 4 章 方差分析	107
4.1 方差分析基本原理和步骤	107
4.2 单因素完全随机设计方差分析	115
4.3 单因素随机区组设计方差分析	124
4.4 系统分组(巢式)设计	127
4.5 二因素(组内无重复)完全随机设计	129
4.6 二因素完全随机设计	131
4.7 二因素随机区组设计	134
4.8 平衡不完全区组设计试验	138

4.9	多因素试验设计	144
4.10	裂区试验设计	149
4.11	重复测量资料方差分析	159
4.12	拉丁方设计	167
4.13	随机区组试验的协方差分析	168
	参考文献	173
第5章	一般线性模型	176
5.1	线性模型基本原理	177
5.2	GLM 模型用户操作界面	182
5.3	GLM 模型输出结果分析	187
5.4	一般方差分析的 GLM 模型	190
5.5	混合效应模型方差分析	192
5.6	系统分组(或嵌套)设计	195
5.7	裂区试验统计分析	197
5.8	协方差分析	199
5.9	数量化方法	201
	参考文献	205
第6章	相关与回归分析	206
6.1	回归和相关概念	206
6.2	两变量的相关分析	207
6.3	直线回归	210
6.4	曲线回归(I: 直线化方法)	219
6.5	曲线回归(II: 非线性最小二乘法)	227
6.6	Deming 回归	237
6.7	局部加权散点光滑(LOWESS)估计	241
6.8	重复观测(试验)回归分析	243
6.9	两随机变量的回归分析	250
6.10	Joinpoint 回归(折线回归)分析	256
	参考文献	263
第7章	分类数据列联表分析	265
7.1	列联表分析及卡方检验概述	265
7.2	列联表的生成与分析	267
7.3	四格表分析	270
7.4	多层 2×2 表 Mantel-Haenszel 检验	275
7.5	$R \times C$ 列联表卡方检验	277

7.6	单向有序 $R \times C$ 表统计检验	279
7.7	双向有序且属性不同的 $R \times C$ 表统计检验	282
7.8	McNemar 检验及 Kappa 检验	287
7.9	$2 \times C$ 表和多层 $2 \times C$ 表	290
7.10	配对病例——对照列联表分析	294
7.11	重复测定资料似然比卡方检验	298
7.12	Poisson 分布抽样情况下多样本检验	300
7.13	列联表行列指标关联的度量	302
7.14	多层列联表 CMH 检验	312
	参考文献	321
第 8 章	分类数据模型分析	323
8.1	分类/计数数据回归模型建模	323
8.2	分类数据回归模型分析用户界面	326
8.3	两分类数据的回归模型	329
8.4	多分类无序反应变量 Logistic 回归	338
8.5	多分类有序 Logistic/Probit 回归模型	341
8.6	Poisson 回归模型	345
8.7	负二项分布回归模型	358
8.8	条件 Logistic 回归	360
8.9	对数线性模型	363
8.10	其他分类数据建模方法	375
	参考文献	384
第 9 章	非参数检验	385
9.1	两样本配对符号检验	386
9.2	两样本配对 Wilcoxon 符号-秩检验	387
9.3	两样本 Wilcoxon 检验	388
9.4	Kruskal-Wallis 检验	390
9.5	中位数检验	393
9.6	Jonckheere-Terpstra 检验	394
9.7	Friedman 检验	395
9.8	Kendall 协同系数检验	398
9.9	Kendall 趋势检验	399
9.10	Cochran 检验	407
9.11	非参数回归分析	408
	参考文献	413

第 10 章 广义线性模型	415
10.1 广义线性模型建模分析基本原理	416
10.2 广义线性模型中的统计检验	423
10.3 广义线性模型拟合程度评价	425
10.4 广义线性模型中的残差诊断	428
10.5 应用广义线性模型统计建模的一般原则	431
10.6 广义线性模型的建模操作	434
10.7 连续变量的广义线性模型	444
10.8 二项分布模型	460
10.9 多项分布的广义线性模型	477
10.10 Poisson 模型	489
10.11 负二项分布模型	502
10.12 Tweedie 分布模型	509
参考文献	514
第 11 章 多因子优化设计与分析	516
11.1 正交试验统计分析	516
11.2 二次正交回归组合(中心复合)设计	526
11.3 Box-Behnken 设计	529
11.4 均匀试验设计	530
11.5 二次饱和 D-最优设计	537
11.6 多因子设计优化分析	538
11.7 含有区组设计的多因子优化分析	549
11.8 “3414”测土配方施肥实验统计分析	554
参考文献	559
第 12 章 混料试验设计与分析	560
12.1 单纯形格子设计	560
12.2 单纯形重心设计	562
12.3 有下界约束条件限制的混料试验设计	562
12.4 单纯形格子设计和单纯形重心设计操作示例	564
12.5 具上下限约束的极端顶点设计	564
12.6 基于均匀设计表的混料试验设计	571
12.7 有上下限条件约束的混料试验设计	573
12.8 具附加线性约束的混料试验设计	578
12.9 混料试验分析的数学模型	580
12.10 混料试验数据的回归分析	591

12.11 偏最小二乘回归分析	595
参考文献.....	599
第 13 章 抽样技术	600
13.1 简单随机抽样.....	600
13.2 分层随机抽样.....	605
13.3 整群抽样.....	615
13.4 系统抽样.....	622
13.5 序贯抽样.....	628
13.6 敏感性问题抽样.....	633
参考文献.....	644

第一篇 DPS[®]数据处理系统

DPS[®]是 data processing system 取首字母缩写而成。该系统采用多级下拉式菜单,用户使用时整个屏幕犹如一张工作平台,随意调整,自如操作,故形象地称其为 DPS 数据处理工作平台,简称 DPS 平台。

DPS 平台将实验设计、统计分析、数值计算、模型模拟以及数据挖掘等功能融为一体,提供了全方位的数据处理功能。

(1) DPS 统计软件是目前国内唯一一款实验设计及统计分析功能齐全、价格上适合于国内用户的具自主知识产权且技术上达到国际先进水平的国产多功能统计分析软件包,并在某些方面已处于国际领先水平。如 DPS 提供的、颠覆了多属性决策领域科学家认知的最大熵-最小残差综合评价数学模型。另外,在均匀试验设计和混料试验设计中采用了独创算法,实现了大型均匀设计表构造和不同类型混料试验设计表构造的重大突破。DPS 数据处理系统亦可确保数据安全、替换国外统计分析软件,如 SPSS、SAS、Minitab 等纯国产统计分析软件。

(2) 完善的统计分析功能涵盖了几乎所有的统计分析技术,是目前国内统计分析功能最全的软件包。其一般线性模型(GLM)可以处理各种类型实验设计方差分析及多元方差分析,特别是一些用 SPSS、SAS 编程较难实现的,如多因素裂区混杂设计、格子设计等方差分析,广义线性模型(GLMs)功能,在 DPS 中变得直观、明了,极易掌握。

(3) 独特、方便的非线性回归建模技术实现了“可想即可得”的用户需求,且参数拟合精度高。

(4) 不断丰富专业统计分析模块适合各个学科的特殊需求,如先后增加了数据包络分析、随机前沿面分析、结合分析、顾客满意度指数(结构方程)模型、数学生态学方法、生物测定、地理统计、遗传育种、生存分析、水文频率分析、量表分析、质量控制图、ROC 曲线分析等内容,并且还在不断地扩充。

(5) DPS=Excel+SPSS。它既有 Excel 那样方便地在工作表里面处理基础统计分析的功能,又实现了 SPSS 高级统计分析技术。DPS 提供十分方便的可视化操作界面,可借助图形处理的数据建模功能为处理复杂模型提供最直观的途径。这些功能是同类软件中所欠缺的。

(6) 有些不是统计分析的功能,如模糊数学方法、灰色系统方法、各种类型的线性规划、非线性规划、层次分析法、BP 神经网络、径向基函数(RBF)、投影寻踪回

归和分类等,在 DPS 里面也都可以找到。

(7) 根据用户要求,不断吸纳新的统计方法,如独立分量分析、小波分析、稳健回归、主导分析、偏最小二乘回归、投影寻踪回归、灰色系统方法、混合分布参数估计、随机森林、面板数据分析、含定性变量的多元逐步回归分析等,并在不断探索,吸纳更多功能,使系统更加完善。

DPS 系统操作简便,在统计分析和模型模拟方面功能齐全,易于掌握,尤其是对广大国内用户来说,其工作界面友好,只需熟悉它的一般操作规则就可灵活应用。

DPS 数据处理系统的研制始于 1988 年,第一版于 1997 年出版发行,当前版本已升级到 DPS 21.05,其运行环境是当前流行的中文 Windows 98/2000/XP/Vista 及 64 位的 Win7、Win8 及 Win10 视窗系统,以及各种国产操作系统。DPS 数据处理系统可广泛适用于教学、科研和生产等各个领域。不管是青年学生还是高级科研人员,不管是计算机应用的初学者,还是经验丰富的计算机应用专家,都可以在本系统中找到自己感兴趣或有用的部分。

本篇介绍 DPS 平台的主要性能和技术特点。

第 1 章 DPS 系统简介

1.1 系统功能简介

1.1.1 独特优良的人机用户界面

一旦开机进入 DPS 平台系统,用户的整个数据处理过程,无论是计算还是数学公式分析都将处于全屏幕编辑状态。可把整个系统视为一张稿纸,键盘和鼠标视为笔,用户可以任意地在上面对写、擦、画表和进行数值计算、统计分析、建立数学模型等,计算结果也都返回到编辑状态下。用户需要做的事情主要是输入编辑数据、定义数据块和在菜单方式下调用统计计算功能。执行统计分析、建立数学模型等高级功能的调用时,只需要用鼠标点一下下拉式菜单,在菜单方式下就可选择执行相应的功能。当相应功能执行完毕后,计算结果存放在编辑器里(当前光标下面,或在另一页中),系统又处于编辑状态。因此用户的任何工作都能在编辑方式下完成。

当进入下拉式菜单操作时,菜单上部是主栏条形菜单,它始终显示在屏幕的顶行,可通过移动鼠标或左右光标键选择主菜单的功能。如从主菜单上选择一项后,将有两种情况可能出现:一是直接转去执行某一项操作程序;二是在主菜单上作出的一项选择后,一个下拉式子菜单就会立即弹出。在子菜单中,使用上下光标键使某项选择变为高亮度,然后再按回车键就可激活该选择项。下拉式子菜单又可产生次一级的下拉式子菜单,当选定相应功能后,按下回车键即可执行。

性能良好的人机用户接口是 DPS 系统特有的操作界面,也是目前国内外同类软件所不具备的。

1.1.2 DPS 已成为国产操作系统下试验数据统计分析工具的不二选择

近年来,为了满足国内对信息安全和自主可控的需求,当前基于 X86、arm 及龙芯架构等的国产化操作系统应用越来越普及,这就需要 DPS 数据处理系统能够适应在不同的操作系统上运行。

为支持 Windows、Linux 等多种操作系统可以使用,DPS 数据处理系统已完成了升级。从过去只能在 Windows 下使用,升级到目前可以在所有国产操作系统上都可使用。随着 DPS 数据处理系统不断升级,DPS 数据处理系统已成为目前国产操作系统环境下试验数据统计分析工具的不二选择。

1.1.3 功能完备的数据编辑电子表格

系统采用目前流行的电子表格作为数据编辑器,本编辑器是针对数据处理的特点而开发的数据处理集成软件。它将文字处理、数据加工和画线制表融为一体,在编辑器的任何地方可随时写入算式或查询有关函数,进行计算或查询函数值。如键入“ $=4.5^3$ ”,即可得到数值“91.125”;键入“ $=ftest(5,20,0.05)$ ”,即可得到方差分析中大均方自由度为 5、小均方自由度为 20、显著水平为 0.05 时的 F 检验临界值 2.7109。

1.1.4 接近自然语言的数学表达方式

DPS 系统在建立数学模型、进行数学模型模拟以及数学计算等方面,其数学表达式非常接近自然形式(甚至公式中的乘号都可省略)。例如,我们输入编辑了两列数,将这两列数据用鼠标选中之后,系统即默认数据矩阵块中第 1 列变量为 x_1 ,第 2 列变量为 x_2 。若第 1 列为因变量,第 2 列为自变量,要建立两变量之间的 Logistic 方程,那么有三个参数需求出。在本系统中,待求参数用 $c_1, c_2, \dots, c_m$ (若有 m 个参数的话)表示,这时只要在下部的文本编辑器中编辑、定义 Logistic 方程的公式“ $x1=c1/(1+\exp(c2+c3x2))$ ”,再选择建立模型的菜单功能项即可得到参数精确估计值、参数标准误差、 t 检验值和置信度等统计参数。

1.1.5 完整实用的多功能统计分析功能

当前版本 DPS 系统的统计分析功能比过去版本更加齐全,根据读者(用户)的建议,增加了若干新的功能,基本上形成了一套较为完整的统计分析体系。

1.1.6 用途广泛的数学模型建模工具

DPS 平台除具有功能齐全的统计分析功能外,还为用户提供了通用的实验数据分析 and 建立数学模型的工具。用户可根据自己的需要,建立符合自己要求和目标而定义的模型。拟建模型可以是单因变量的线性回归模型或非线性经验模型,也可是由多个因变量建立的联立线性回归模型或非线性模型方程组。这些线性、非线性模型又可以是差分模型,或是对各个变量进行加权处理后的模型。

此外,DPS 系统还可对模型进行模拟分析。例如,数学模型中各个参数的变化速率分析(如农业技术经济学中的边际分析)、求解模型的最大值和最小值(即建立模型后对模型模拟寻优)以及线性规划、非线性规划等,DPS 系统均有相应的功能模块。

用 DPS 建立和分析数学模型,可使复杂的工作变得简单易行,用户只需按系统规则定义模型就可以了。DPS 系统在数学模型处理方面实现了“所想即所见,所见即所得”。

1.1.7 简便的数值计算

方程求解包括求解线性规划或非线性方程组(如无约束非线性规划)。

DPS 系统对矩阵的处理采用交互方式,用图标实现了矩阵计算功能,包括两个矩阵相加、相减、相乘(左乘或右乘)、矩阵求逆、矩阵转置、求矩阵特征值及特征向量等繁琐的计算工作。其他数学分析和计算功能包括实系数多项式求根、微分方程求导数值、定积分、多重积分数值的计算以及微分方程的初值求解。

DPS 系统的数学计算功能,特别是矩阵交互式处理功能使得当前高等院校在数理统计分析教学工作中遇到的原本繁冗无味的矩阵计算变得快速而精确,可以在很大程度上提高教学质量,起到很好的辅助教学作用。

1.1.8 丰富多样的数学函数

DPS 系统定义的数学函数除包括一般常用数学函数,如 sqrt(求平方根)、exp(以 e 为底的幂)、lg(以 10 为底的常用对数)、ln(以 e 为底的自然对数)外,还提供了数理统计分析中常用的统计函数。主要包括以下类群:

(1) 数理统计检验临界值,包括 Fisher 检验临界值、Student 的 t 检验临界值、卡方(χ^2)检验临界值、相关系数检验临界值等。

(2) 数理统计量置信度,如 Fisher 检验置信度、Student 的 t 检验置信度、卡方检验置信度以及相关系数检验置信度等;概率分布函数,如 F 分布的分布函数、卡方分布的分布函数、 t 分布的分布函数、二项分布的分布函数及 Poisson 分布的分布函数等。

此外,DPS 系统还提供了计算生物群落多样性指数等多种参数的计算功能。

1.1.9 直观捷达的数据作图辅助分析系统

在数据资料比较及动态分析中,图形以其直观的表达深受人们欢迎,但不足之处是处理速度慢,费时费力。而 DPS 系统的另一突出特点就是将数据分析和数据作图有机地结合起来,极大程度地弥补了这一缺憾。有些分析结果系统除了以文字形式输出外,还自动以图形方式给出(如聚类谱系图),当然也可以对分析数据作图,如拟合曲线图、 x - y 折线图等。

1.1.10 汉英双语的用户界面

DPS 系统同时提供给用户中文版和英文版。中英文版本的用户菜单提供了一键切换的汉英双语界面。它是不可多得的数据统计分析方面的汉英双语词典,并可使 DPS 系统适用于不同类型的用户群。

1.2 DPS 系统的不同版本

为适应不同用户群的需求,DPS 采用不同版本形式,即标准版、高级版和多用户高级版。

标准版采用注册机制,一机一号,提供给用户使用。用户安装后,将机器号发送到软件开发者,缴费后,即可得到一个注册号。填入注册号即可使用。标准版在统计分析功能方面不再升级,其数据编辑处理的电子表格最大 65535 行 $\times$ 255 列。

高级版用户配一个软件狗可以在多台计算机上使用。无论是 Windows 操作系统还是国产操作系统,只要在使用时插上软件狗即可。高级版将根据用户需求不断升级。用户升级是免费的。其数据编辑处理的电子表格最大可达数百万行、数百万列。

多用户高级版可以同时多台机器上安装使用。和高级版一样,多用户高级版将根据用户需求不断升级。用户升级是免费的。其数据编辑处理的电子表格最大也可达 20 万行 $\times$ 20 万列。

本章其他各节主要介绍 DPS 系统高级版的数据编辑处理操作。

1.3 系统运行环境与安装、使用

1.3.1 Windows 操作系统下安装与运行

Windows 98/2000/XP/Vista/Win7/Win8 以及 Win10 操作系统,硬盘空间至少 500MB,显示方式推荐 1024 $\times$ 768,VGA 显示,内存至少 512MB,最好是 2GB。

DPS 的安装:①从 DPS 数据处理系统官方网站 [http:// www. dpsw. cn](http://www.dpsw.cn) 的“下载中心”下载 DPS 数据处理系统高级版(最新版本)或标准版,保存在桌面或其他地方。②在桌面或其他地方找到下载的 DPS 安装程序:DPS 高级版的文件名为“dps_ENT. exe”,标准版的文件名为“dps_2006. exe”。③用鼠标双击 DPS 的安装程序,打开 DPS 安装程序,进行安装。④在安装过程中,根据安装提示,选择系统路径和其余参数。一般情况下,只要单击“下一步”按钮即可。如果是在 Windows 64 位系统下安装 DPS,则需要在如下的安装界面中选中“Windows 64 位系统”选项(图 1-1)。⑤安装完毕时,系统有时需要重新启动。重新启动后,系统会自动在桌面上建立一个“DPS 数据处理系统”图标。

1.3.2 在 64 位 Windows 下 DPS 应用程序属性设置

如果是在 Win7/Win8/Win10 的 64 位系统下运行 DPS 软件,在运行之前建议设置 DPS 应用程序的属性。即用鼠标点击 DPS 数据处理系统图标,然后按下鼠标右按钮,再用鼠标点击属性,出现如图 1-2 所示的用户界面。用鼠标点击界面

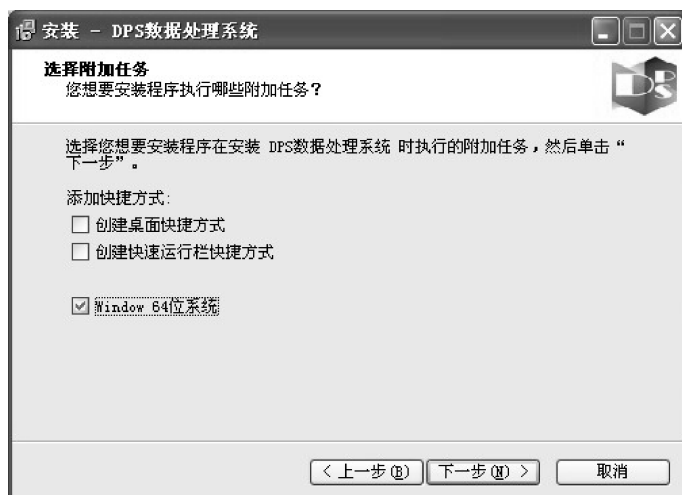


图 1-1 Windows 64 位系统安装时选项设置

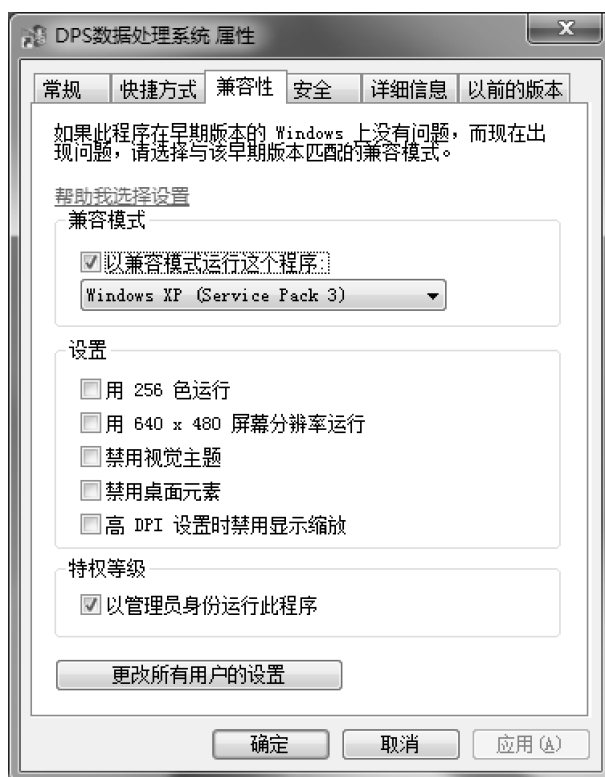


图 1-2 Windows 64 位操作系统下应用程序属性设置

中的“兼容性”项,并将里面的“以兼容模式运行这个程序”下面的选项选“Windows XP”,并打勾;再在“特权等级”下面的“以管理员身份运行此程序”的前面打勾。不然,在使用 DPS 时,有的功能,如非线性回归模型参数估计等会无法运行。

1.3.3 DPS 应用程序的启动与退出

在桌面上双击 DPS 图标即可进入 DPS 数据处理系统。

如果用户是从网站下载的 9.50 标准版的试用版,系统会出现提示用户进行注册的用户界面。没有注册的 DPS 试用版,系统对用户的使用权限有限制。一旦申请注册成功,DPS 将出现图 1-3 所示的用户界面。



图 1-3 DPS 系统用户注册成功后的界面

当用户数据处理、统计分析结束时,可在菜单下的“文件”选单中,选择“退出”命令,或用键盘输入“Alt+F4”指令结束操作。这时系统会提示用户是否保存当前文件内容,如图 1-4 所示。

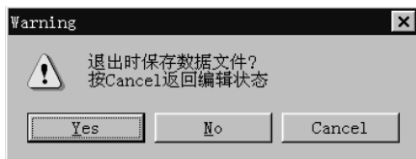


图 1-4 系统退出时是否保存文件界面

这时按“Yes”按钮将保存文件并退出,按“No”按钮将不保存文件并退出,按“Cancel”按钮将不退出,返回编辑状态。

1.4 国产操作系统下 DPS 数据处理系统安装与使用

使用 DPS 数据处理系统,首先需按照国产电脑的系统架构类型进行下载,如

常用的兆芯、海光和飞腾 CPU,按对应的链接下载。如果下载错了,则安装不上。首先在官网上找到适合你的系统架构(如 X86 或龙芯)的安装包并下载。注意下载的安装包格式,如. deb 等。



然后在国产操作系统下,点击安装包文件,按照系统提示进行安装。国产电脑系统安装软件的方式因操作系统类型而异,但通常都提供了较为简便的安装途径。以下以统信 UOS 为例,介绍国产电脑系统安装软件的一般步骤:

安装应用软件前,建议先更新系统,确保系统的安全性和稳定性。安装完成后,建议检查应用软件的设置和权限,确保个人隐私和系统安全。

当我们找到自己需要安装的软件时,一般情况下只需轻轻一点,即可开始安装过程,不过,需要注意的是,如果有的国产操作系统尚未激活,可能无法进行安装。因此,在安装应用之前,确保电脑系统已经完成激活是十分必要的。

对于熟悉命令行的用户,还可以通过命令行来安装应用软件。具体的命令可能因软件和应用而异,但通常涉及使用“dpkg”或“apt”等命令来安装软件包。

1.5 DPS 数据处理系统用户工作界面

在桌面上双击 DPS 的图标即可进入 DPS 数据处理系统编辑状态。进入编辑状态后,系统会出现如图 1-5 所示的用户工作界面。

按图 1-5 从上到下的顺序,DPS 工作界面包含如下几项内容:标题栏、菜单栏、工具栏、下拉菜单、工作表、当前工作表、活动单元格、工作表标签及文本编辑区,下面分别介绍它们的作用。

1. 标题栏

标题栏告诉用户正在运行的程序名称和正在打开的文件名称。如图 1-5 所示,标题栏显示此窗口的应用程序为 DPS,在 DPS 中打开的当前文件的文件名为“dps1.dps”。

2. 菜单栏

菜单栏按功能把 DPS 命令分成不同的菜单组,它们分别是文件、数据编辑、数

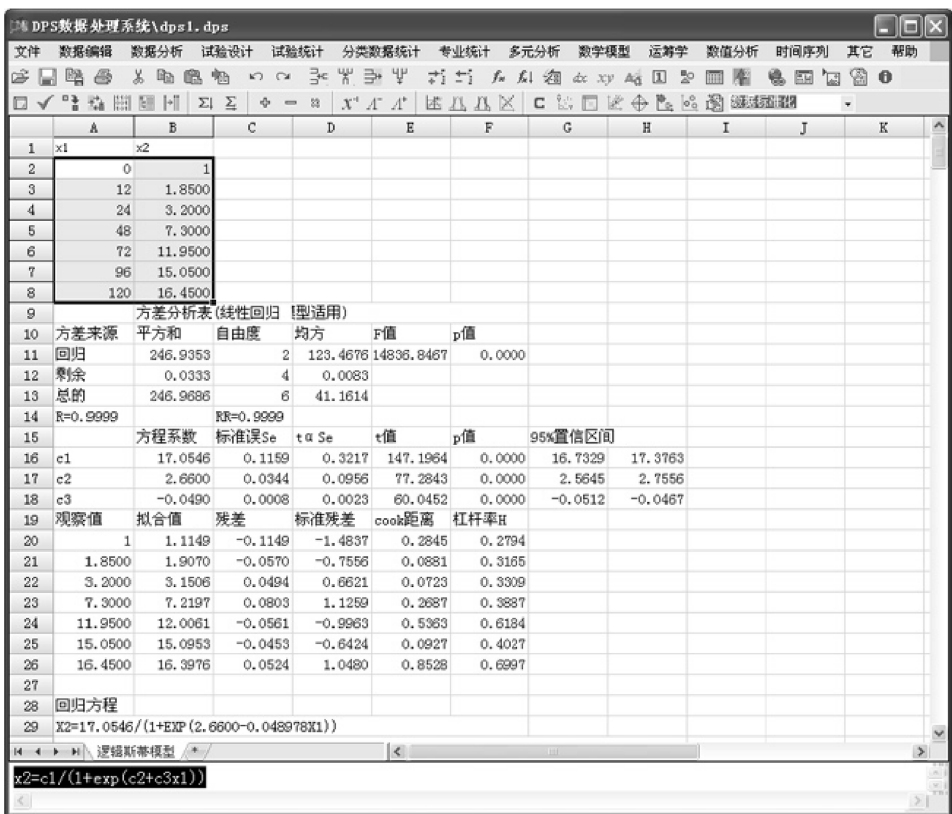


图 1-5 DPS 用户工作界面

据分析、试验设计、试验统计、分类数据统计、专业统计、多元分析、数学模型、运筹学、数值计算、时间序列及其他等。当菜单项被选中时,引出一个下拉式菜单,可以从中选取相应的子菜单。另外,在屏幕的不同地方单击鼠标右键时,“快捷菜单”将出现在鼠标指针处。选取“快捷菜单”中的命令与从菜单栏的菜单上选取相应命令的效果是一样的,但选取速度明显加快。

3. 工具栏

DPS 提供了几种工具栏,这些工具制简化用户的操作。工具栏中的按钮都是常用命令,当鼠标指向某一按钮后,稍等片刻在按钮右下方就会显示该按钮命令的含义。

工具栏的第一行是常用工具栏。常用工具栏中为用户准备了使用 DPS 编辑、处理数据最常用命令的快捷按钮,如“新建文件”按钮、“打开文件”按钮、“保存文件”按钮。“格式”按钮处理那些和文本外观有关的命令,如字体、字号、对齐方式及